IDEA PROJECT FINAL REPORT
Contract ITS-17

IDEA Program
Transportation Research Board
National Research Council

October 1, 1995

A SEQUENTIAL HYPOTHESIS TESTING-BASED DECISION-MAKING SYSTEM FOR FREEWAY INCIDENT RESPONSE

Samer Madanat and Hualiang Teng
Purdue University

The ITS-IDEA program is jointly funded by the U.S. Department of Transportation's Federal Highway Administration, National Highway Traffic Safety Administration, and Federal Railroad Administration. For information on the IDEA Program contact Dr. K. Thirumalai, IDEA Program Manager, Transportation Research Board, 2101 Constitution Avenue N.W., Washington, DC 20418 (phone 202-334-3568, fax 202-334-3471).

INNOVATIONS DESERVING EXPLORATORY ANALYSIS (IDEA) PROGRAMS MANAGED BY THE TRANSPORTATION RESEARCH BOARD (TRB)

This investigation was completed as part of the ITS-IDEA Program which is one of three IDEA programs managed by the Transportation Research Board (TRB) to foster innovations in surface transportation. It focuses on products and result for the development and deployment of intelligent transportation systems (ITS), in support of the U.S. Department of Transportation's national ITS program plan. The other two IDEA programs areas are Transit-IDEA, which focuses on products and results for transit practice in support of the Transit Cooperative Research Program (TCRP), and NCHRP-IDEA, which focuses on products and results for highway construction, operation, and maintenance in support of the National Cooperative Highway Research Program (NCHRP). The three IDEA program areas are integrated to achieve the development and testing of nontraditional and innovative concepts, methods and technologies, including conversion technologies from the defense, aerospace, computer, and communication sectors that are new to highway, transit, intelligent, and intermodal surface transportation systems.

Table of Contents

A SEQUENTIAL HYPOTHESIS TESTING-BASED DECISION-MAKING SYSTEM FOR FREEWAY INCIDENT RESPONSE

1. Executive Summary

A decision-making system for integrated incident detection and response has been developed. The system employs sequential hypothesis testing techniques to dynamically optimize incident response decisions. It systematically considers the trade-offs between the possible costs of a delayed incident response decision and the improved decision-making capabilities that result from delaying action until additional measurements are taken. In the first stage of the research, the mathematical model underlying the decision-making system was developed and the solution algorithm was implemented using a rolling-horizon framework. In the second stage, the decision-making system (DMS) was evaluated using simulation testing, for a variety of traffic scenarios. For each traffic condition, the performance of the proposed system was compared to that of a state-of-the-art incident detection algorithm. It was found that the DMS achieved substantially shorter incident response time without increasing the false-response or the non-response rates.

2. Body

IDEA product

The product from this research is a decision-making system for integrated freeway traffic incident detection and response. This system employs sequential hypothesis testing techniques to dynamically optimize incident response decisions by systematically considering the tradeoffs between the possible costs of a delayed incident response decision and the improved decision-making capabilities which result from delaying action until additional measurements are taken.

The input components to this model include traffic parameters, their distributions under different traffic conditions, traffic delay costs due to incidents, the costs of implementing response measures, incident frequencies in time and space, and the distribution of the incident durations. The outputs are optimal incident response policies for each time period. In real-time operations, the derived optimal policies can be used to select incident response decisions, given various traffic conditions.

Concept and innovation

The proposed system is based on a novel approach to the incident response process. This decision process is modeled as an optimization problem in which the uncertainties in the measured traffic stream characteristics and the costs associated with incorrect decisions are considered simultaneously.

Because incident detection and response decisions are made simultaneously, this system can be viewed as an incident detection system. Compared to conventional incident detection systems, the proposed system explicitly accounts for the presence of traffic stream measurement and interpretation errors, and simultaneously considers incident detection decisions and possible response actions such as dispatching emergency vehicles and traffic diversion to alternative routes.

Investigation

Our example traffic network used in the model formulation is illustrated in Figure 1. The simple network consists of a freeway link and a uni-directional surface street. A pretimed traffic signal controls the junction of the on-ramp and surface street, while the junction of the off-ramp and street is uncontrolled. The only freeway traffic management strategy used in this study is route diversion to the surface street via the dissemination of incident information to freeway users.

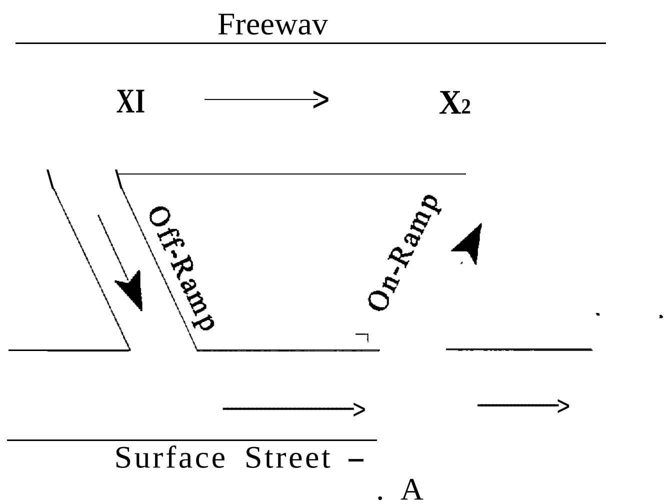

FIGURE 1 Hypothetical Traffic Network

The incident response decision-making process is formulated as a sequential hypothesis testing problem. In each time period, the traffic control system obtains measurements of freeway traffic conditions from a surveillance system and thus is confronted with two mutually exclusive hypotheses, defined as:

H_0: no incident has occurred on the freeway, and

H_1: an incident has occurred on the freeway.

After each observation, the decision-making system will either:

(1) accept H_0 and implement no response,

(2) accept H_1 and initiate route diversion from the freeway to the surface street, or

(3) delay the decision to accept either hypothesis for an additional measurement period.

The decision ofwhether to accept either hypothesis or to delay the acceptance is based on the expected losses associated with these decisions. The expected losses are computed on the basis of the current non-incident probability. This probability is a function of all previously measured traffic conditions and of the probability distributions of these measurements under incident and non-incident situations.

Mathematically, the current non-incident probability can be obtained recursively by Bayesian updating as follows:

$$P_t = \frac{p_{t-1} f_0(z_t)}{P_{t-1} f_0(z_t) + (1-p_{t-1}) f_1(z_t)} \quad , t=1,2,... \qquad (1)$$

where: p_t: non-incident probability at time t;

$f_0(z_t)$, $f_1(z_t)$: probability density functions of the traffic measurements for non-incident and incident conditions, respectively;

z_t: traffic measurements obtained at the beginning of period t; examples of traffic measurements include occupancy, speed and traffic flow.

p_0: the prior used in the Bayesian updating formula which is obtained by a set of incident likelihood prediction models as described in Appendix A.

The probability density function $f_0(z_t)$ and $f_1(z_t)$ must be calibrated from field data for specific locations. Figure 2 shows the occupancy probability density functions for detector data obtained from a Toronto freeway (1).

The optimal response policy can be solved using dynamic programming (2). The dynamic programming formulation is given by:

$$J_t(p_t) = \min \{(1-p_t)L_0,\ p_t L_1,\ (1-p_t)C + \underset{z_{t+1}}{E}[J_{t+1}(p_{t+1})]\}, \quad t=1,2,... \qquad (2)$$

where:

$J_t(p_t)$: minimum expected cost-to-go function for time period t and state variable p_t;

L_0: loss resulting when no response is made to an incident; given by the expected difference between the delay incurred with and without diversion.

L_1: loss associated with a false response; given by the expected total travel time increase due to unwarranted traffic diversion from the freeway to the surface street.

C: cost associated with waiting one more period before making a response, if an incident has indeed occurred on the freeway. This cost is incurred by the drivers who pass location X_2 in Figure 1 during one measurement interval.

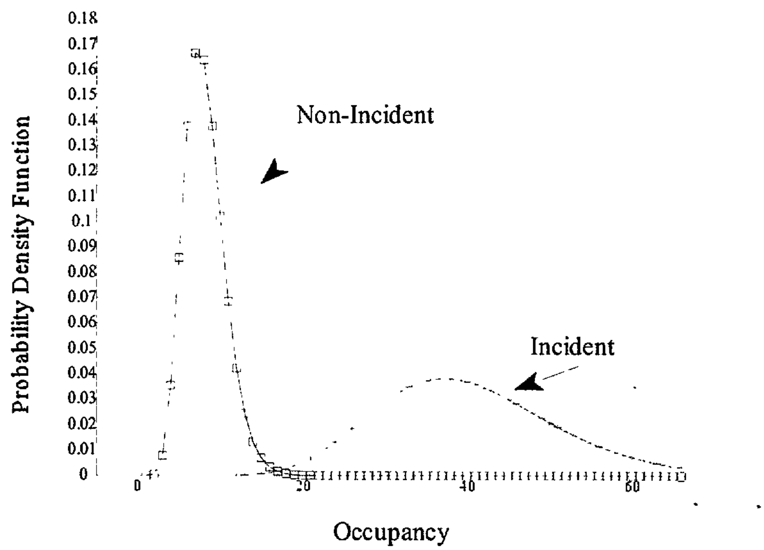

FIGURE 2 Incident and Non-incident Probability Density Functions for Occupancy Measurements

The optimal policy can be shown to be stationary if the predicted costs L_0, L_1 and C are constant and deterministic, and if the analysis period is sufficiently long. This optimal policy is defined by a closed-form expression (2) which has the desirable property of minimal computational complexity, a benefit in the context of on-line traffic management. To exploit this property, a rolling-horizon implementation of the decision-making algorithm was used.

The rolling horizon implementation is a certainty-equivalent control. At the beginning of each control interval, short-term predictions of traffic conditions are used to derive constant and deterministic costs-per-stage (L_0, L_1 and C). These costs are used in the dynamic programming algorithm to derive a stationary optimal policy for a predetermined projection horizon. This projection horizon rolls forward at the end of each control interval, as shown in Figure 3, and the cost functions are updated on the basis of the latest traffic measurements. Thus, at the beginning of each projection horizon, a new incident response policy is derived for that horizon, but applied for the first control interval only.

For the projection horizon beginning with control interval τ, the dynamic programming formulation is given by:

$$J_t(p_t) = \min \{(1-p_t)L_0^\tau,\ p_t L_1^\tau,\ (1-p_t)C^\tau + \underset{z_{t+1}}{E}[J_{t+1}(p_{t+1})]\} \quad, t=\tau,\tau+1,\ldots,\tau+H-1, \qquad (3)$$

where the costs-per-stage are superscripted by the index of the control interval τ, and H is the length of the projection horizon.

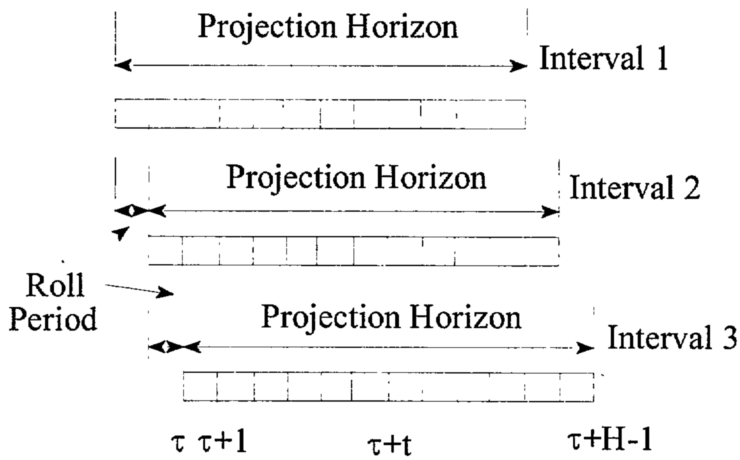

FIGURE 3 Rolling Horizon Implementation

The following equalities are assumed to hold for each projection horizon:

$$L_0^t = L_0^\tau \text{ , t=}\tau,\tau+1,\ldots\tau+H-1, \tag{4}$$

$$L_1^t = L_1^\tau \text{ , t=}\tau,\tau+1,\ldots\tau+H-1, \text{ and} \tag{5}$$

$$(1-p_t)C^t = (1-p_\tau)C^\tau \text{ , t=}\tau,\tau+1,\ldots\tau+H-1, \tag{6}$$

The stationary optimal policy for control interval τ takes the form (2):

accept H_0, if $p_t \geq \alpha^\tau = 1 - \dfrac{(1-p_\tau)C^\tau}{L_0^\tau}$

accept H_1, if $p_t \leq \beta^\tau = \dfrac{(1-p_\tau)C^\tau}{L_1^\tau}$

collect one more observation, if $\beta^\tau < p_t < \alpha^\tau$. (7)

This stationary policy is known as the Sequential Probability Ratio Test (SPRT).

The operation of the decision-making algorithm is illustrated in Figure 4. At the beginning of each control interval τ, new measurements of traffic flow, occupancy and average speed are obtained. These measurements are used to update the incident probability for the current interval. These measurements are also combined with previous measurements to predict traffic conditions for the remainder of the rolling horizon. These traffic predictions are used to update the cost functions for the current control interval. Given these updated functions and incident probability, the response policy thresholds for this interval are obtained. The optimal response decision for the current interval is obtained by comparing the updated incident probability to the updated thresholds.

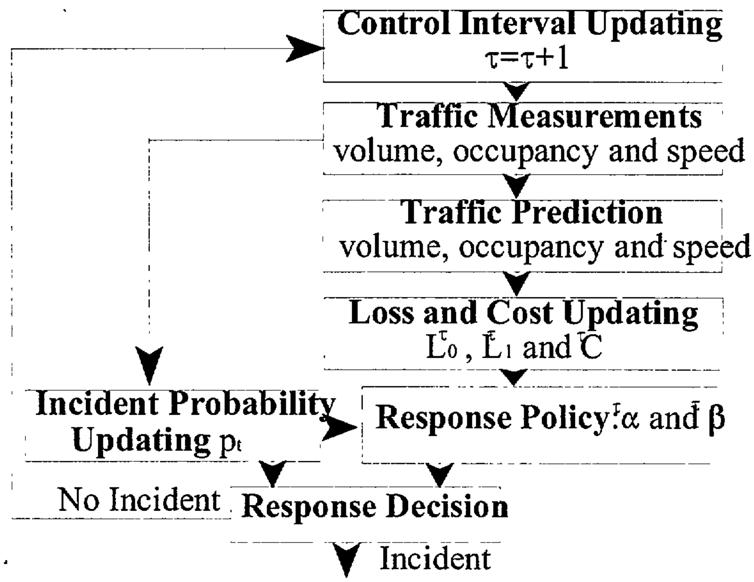

FIGURE 4 Flow Chart of the Incident Response Decision-Making Algorithm

The SPRT algorithm was evaluated using the INTRAS freeway simulation model (3). The evaluation procedure is illustrated in Figure 5. The inputs to INTRAS consist of traffic data such as flow and speed, geometric data and incident related data such as incident type, location and duration. Given these inputs, INTRAS simulates occupancy readings at specified detector locations, in 20 second intervals. These occupancy readings are then used as inputs for the SPRT algorithm. INTRAS also emulates the detection decisions of the California Algorithm 7 (4). The evaluation process involves comparisons of the SPRT and the California algorithms on the basis of three criteria: response time, non-response rate, and false response rate. Two scenarios are simulated: the presence and the absence of an incident. For each scenario, the traffic flow is varied from low (3000 vph) to high (5000 vph), and the incident location is changed from 1800 ft to 2600 ft upstream of the detectors. The diversion rate is varied from 5% to 20%. Some parameters in the cost functions are kept constant for the duration of the evaluation. For example, the diversion period is set to 10 minutes, the mean incident duration is fixed at 15 minutes, and the incident bottleneck capacity is set to 2000 vph. The results of the evaluation are listed in Table 1.

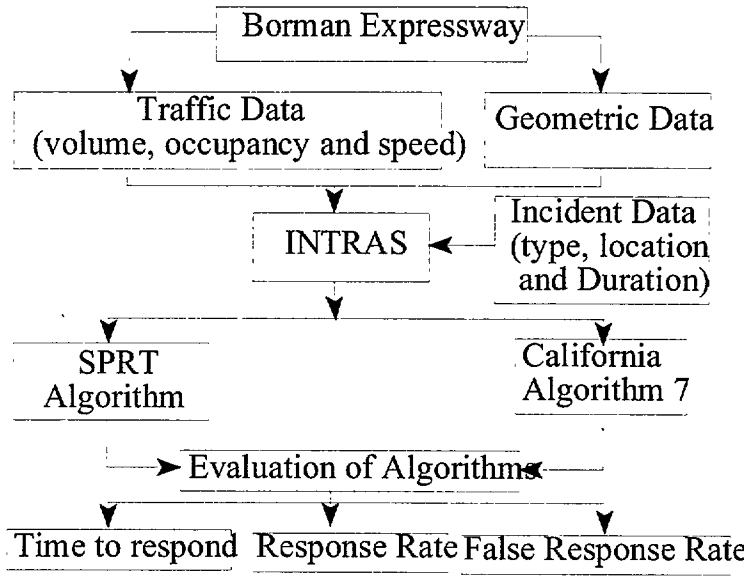

FIGURE 5 Simulation Evaluation Framework

The SPRT algorithm has a significantly shorter incident response time than the California algorithm 7 and the algorithms' performance with respect to non-response rate and false response rate are comparable. Because the SPRT algorithm's non-response rate and false response rate are both zero, and its response time is the shortest, minimum cost is incurred to traffic. Certain factors such as traffic flow and incident location influence the performance of the algorithms. Both algorithms respond slower to an incident which occurs further from the detectors, as the shock wave caused by the incident requires a longer time to reach the loop detector. In high traffic flow conditions, incident response is more rapid than in low flow conditions, as the speed of the backward- moving shockwave created by the incident is greater in high flow conditions.

TABLE 1 Simulation Results

Scenarios		SPRT			California Algorithm		
Volume	Incident Location (ft)	Response Time (seconds)	Non-Response Rate	False Response Rate	Response Time (seconds)	Non-Response Rate	False Response Rate
Low	1800	640	0%	0%	860	0%	0%
	2600	940	0%	0%	1125	0%	0%
High	1800	231	0%	0%	376	0%	0.63%
	2600	336	0%	0%	550	0%	0.63%

Plans for implementation

This section discusses some of the implementation issues for the incident response decision-making system developed in this research. These issue include: off-line data requirements, on-line input requirements and computational needs. To date, our decision-making system was calibrated for on-line inputs received from loop detectors. To accommodate outputs from other types of detectors, the system's models must first be calibrated off-line using historical data. Using multiple detector types within the response algorithm is expected to increase the efficiency of the decision-making system.

Currently, the decision-making system is operational for a single freeway section as described in Figure 1. Therefore, the following discussion will emphasize implementation issues for a single freeway section including the associated on-ramp, off-ramp and corresponding surface street section. The implementation of the current version of the decision-making system requires the presence of three loop detectors: one detector located immediately downstream of the off-ramp, one located immediately upstream of the on-ramp, and one located on the off-ramp. If a single measurement of traffic occupancy is used in the SPRT algorithm, the occupancy data is measured at the first detector. This detector also provides measurements of traffic flow and speed for cost computation and prediction at the beginning of each control interval. The second detector provides measurements of traffic flow and speed downstream of the bottleneck, which are used for cost computation. The traffic volume obtained from the off-ramp detector is used for measuring and updating the traffic diversion rate from the freeway onto the surface street.

The following parameters must be specified off-line, prior to the operation of the decision-making system.

1) The type of traffic measurements used; currently, our decision-making system accepts two types of traffic measurement: upstream detector occupancy or relative spatial occupancy difference between upstream and downstream freeway detectors.

2) The time interval between two observations; in the parametric analysis performed in this study, we use 20 seconds.

3) The prior probability of non-incident; this is the output of the incident likelihood prediction models, described in the Appendix of this report.

4) The length of the freeway section.

5) The capacity of the freeway section.

6) The free flow speed of the freeway section.

7) The length of the surface street section.

8) The uncongested speed on the surface street section.

9) The existing surface street flow.

10) The initial fraction of freeway traffic that diverts in incident conditions; this quantity is updated using the moving average method at the beginning of each control interval, after obtaining new measurements from the off-ramp loop detector.

11) The initial estimate of the length of time period for which diversion from freeway to surface street is performed.

12) The signal cycle length on the surface street.

13) The signal green time on the surface street for the traffic stream traveling in the direction parallel to the freeway section.

The decision-making system uses the following on-line inputs, measured by the three loop detectors, and updated at the start of each control interval:

1) Traffic flow at upstream detector.

2) Average vehicle speed at upstream detector.

3) Upstream detector occupancy.

4) Traffic flow at downstream detector.

5) Average vehicle speed at downstream detector.

6) Downstream detector occupancy.

To extend the decision-making system to a freeway system consisting of multiple sections, each equipped with the three detectors described above, the SPRT algorithm can be applied sequentially one section at a time starting at the downstream end of the freeway system. Referring to Equation (7), the derivation of the incident response decision for each section involves simple algebraic operations, due to the closed form response policy. Solving for the optimal incident response for one section took a fraction of a second, when the algorithm was implemented on a Pentium personal computer during the simulation-testing experiments performed as part of this research project. Therefore, the computation time required for running the algorithm in the case of a freeway system consisting of N sections is less than N seconds. In on-line operations, the optimal response decision for each freeway section can be activated as soon as it has been solved for, and before the policies of the upstream sections have been obtained. This means that the SPRT algorithm can be applied to a freeway system of any length with the use of the standard computational resources available to Traffic Operations Centers, namely stand-alone Pentium PCs or Workstations.

The Indiana Department of Transportation has expressed interest in the product of this research effort. The incident response decision-making system will be incorporated in the Borman Advanced Traffic Management System, after it has been extended to handle multiple sections. The decision-making system can be used by traffic control personnel to assist in responding to various freeway incidents in a near optimal manner, to minimize traffic delays and reduce the number of secondary incidents.

3. Conclusions

In this research, we have developed a new methodology for freeway incident response decision-making. This methodology, which is based on the Sequential Hypothesis Testing framework, explicitly accounts for the losses associated with incorrect detection and response decisions and optimizes the tradeoffs between these expected losses. To facilitate the application of the decision-making methodology within the constraints of on-line traffic management, a rolling-horizon implementation was used. Results obtained by simulation indicate that the new decision-making system achieves shorter incident response times than traditional incident detection algorithms, without increasing the false-alarm and non-response rates. This superior performance can be attributed to the fact that the new system explicitly minimizes the sum of the expected losses associated with the response decisions.

References

1. C.-H. Hsiao. The Application of Fuzzy Logic and Neural Networks to Freeway Incident Detection. PhD Dissertation, School of Civil Engineering, Purdue University, 1994. .

2. D.P. Bertsekas. Dynamic Programming: Deterministic and Stochatic Model. Prentice-Hall, Inc., 1987.

3. D.A. Wicks and E.B. Lieberman E.B. Development and Testing of INTRAS, a Microscopic Freeway Simulation Model, Vol. 1, Program Design, Parameter Calibration and Freeway Dynamics Component Development. Report No. FHWA/RD-80/106, Federal Highway Administration, Washington D.C..

4. H.J. Payne and S.C. Tignor. Freeway Incident-Detection Algorithms Based on Decision Trees with States, Transportation Research Record 682, pp. 30-37, 1978.

5. M. Ben-Akiva and S. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand, the MIT Press, Cambridge, MA, 1985.

Appendix A

A Prototype System for Real-Time Incident Likelihood Prediction (ITS-2)

This appendix describes the development of the incident likelihood prediction models that provide the prior probabilities used in the computation of current incident probabilities as shown in Equation (1) of this report. This work was performed as part of research contract ITS-2, which was awarded to Purdue University by the National Research Council. The work was performed under the supervision of Dr. Samer Madanat.

A.l. Problem Statement

The first objective of this research project was to develop models which can be used to provide real-time predictions of freeway incident likelihoods. Such predictions will serve as the basis for a proactive corridor-wide traffic control system. In such a system, traffic stream and environmental conditions measured by surveillance sensors would be used as inputs for predicting incident likelihoods in near real-time. Traffic control strategies can thus be immediately implemented to reduce the probability of an incident, as well as to mitigate incident-related problems if they occur.

To prove the feasibility of this concept, it was essential to demonstrate the possibility of accurate predictions of freeway incident probabilities, based on near real-time measurements of traffic and weather variables. As described in the following section of this report, we have successfully developed models for likelihood prediction of two critical types of freeway incidents: crashes and overheating vehicles. These models capture the influence of various traffic and weather factors on the probabilities of vehicle crash and overheating vehicle incidents. Furthermore, both models have high internal and external validity, as demonstrated by their fit to the data and their predictive accuracy, respectively.

The predictions given by the incident likelihood models can be combined with measurements obtained by loop detectors to improve the accuracy of the estimates of incident probabilities. State-of-the-art incident detection algorithms utilize only traffic information. By considering both traffic and environmental variables, it is possible to achieve a more accurate estimate of incident probability. This estimate is used as an input to a sequential incident-response decision-making process, as shown in the body of this report.

A.2. Research Approach

This section describes the approach that was used in developing the freeway incident likelihood prediction models. Because the outputs of the incident prediction models are probabilities of a binary event, an appropriate methodology to use is binary logit Binary logit is a powerful tool which has been widely used in transportation demand modeling studies.

Eight-and-a-half months of incident, traffic and weather data for the Borman expressway were used for model development. We sampled non-incident data from the non-incident population which comprises those time periods in which no incidents were observed. Therefore, our sample is a stratified random sample with two strata, incidents and non-incidents.

Two binary logit incident prediction models are presented in the following paragraphs. These are models for two types of incidents: (i) overheating vehicles, and (ii) crashes. In Tables Al & A2 the column entitled "Independent Variable" lists the explanatory variables used in the model. The

“Estimated Coefficient” column shows the contribution of each explanatory variable to the probability of that type of incident and the “t-Statistic” column displays the statistical significance of that variable. A t-statistic larger than 1.65 in absolute value means that the variable is a significant predictor of that type of incident at the 90% confidence level. The goodness of fit of each model is represented by ρ^2; the larger the value of ρ^2, the better the fit of the model to the data. In binary logit models, the statistic “percent correctly predicted” provides an estimate of the predictive accuracy of each model.

For the overheating vehicle incident likelihood model, the variables peak, merge, temp (temperature), rain, and spv (speed variance) were found significant.

The coefficient for the variable peak has a positive sign, which suggests that an overheating vehicle incident is more likely to occur in a peak period than a non-peak period. This is expected because traveling speeds are slower during the peak period. This variable is not significant at the 90% confidence level, as can be seen by the value of its t-statistic (1.62), possibly because the peak period on the Borman expressway is widely spread out. The coefficient of the variable merge represents the effect of location relative to on/off ramps on the likelihood of an overheating vehicle incident. The positive sign of this coefficient indicates that an overheating vehicle incident is more likely to occur in a merge section than a mid-section. The value of the t-statistic(2.19) suggests that this effect is significant. The coefficient of the variable temp shows the effect of temperature on the likelihood of an overheating vehicle incident. The positive sign suggests that an overheating vehicle incident is more likely to occur in high temperature conditions than low temperature conditions, because high temperatures aggravate engine overheating. The high t-statistic (4.63) strongly supports this explanation. The coefficient of the variable rain has a negative sign which indicates that an overheating vehicle incident is more likely to occur in sunny (non-rainy) conditions than in rainy conditions. The t-statistic (-2.29) shows a significant effect for the variable rain. The coefficient of the variable spv represents the effect of speed variance between lanes on the likelihood of an overheating vehicle incident. The negative sign means that an overheating vehicle incident is more likely to occur in lower speed variance conditions than higher speed variance conditions. This is because when the speed variance is low, there are less overtaking opportunities, which increases the likelihood of an overheating vehicle incident. The t-statistic (-2.37) suggests that this result is significant. Overall, this model demonstrates good fit to the data, as can be seen from the value of ρ^2 (0.21), and high predictive accuracy, as measured by the high percentage of observations correctly predicted (74%).

For the crash model, the variables merge, visi (visibility), and rain are found significant. In Table 2, the coefficient of the variable merge has a positive sign, which suggests that a crash is more likely to occur in a merge section than a non-merge section. Though the t-statistic (1.46) indicates that this variable is not strongly significant at the 90% confidence level, it has the correct sign, because there are more vehicle interactions and therefore a higher crash probability in the merge sections, where traffic flow is not as smooth as in the mid-sections. The coefficient of the variable visi has a negative sign, which indicates that a crash is more likely to occur in low visibility conditions, as expected. This variable is not strongly significant, as can be seen by its t-statistic (-1.02) possibly because, in our dataset, visibility is measured in miles, a unit which is not sufficiently precise to capture the effect of low visibility on drivers. The coefficient of the variable rain has a positive sign, which means that a crash is more likely to occur in rainy conditions than non-rainy conditions. This is because the presence of rain reduces visibility and lowers pavement skid resistance. The high t-statistic (3.45) supports this explanation. The fit of this model is satisfactory, as shown by its ρ^2 value (0.14), as is its predictive accuracy (71% of observations correctly predicted).

It should be noted that the estimated coefficients in these models are unbiased regardless of the use of a stratified random sampling scheme in which incidents are oversampled. The only correction that must be made is for the constant, using the method described in (5). The effect of this correction is to reduce the probability of an incident by a factor proportional to the log of the fraction of incident observations in the sample divided by the fraction of incident observations in the population.

Table Al Incident Likelihood Model 1 for Overheating Vehicles

Independent Variable	Estimated Coefficient	t-Statistic
constant	-2.45	-5.25
peak	0.40	1.62
merge	0.51	2.19
temp	0.03	4.63
rain	-1.06	-2.29
spv	-0.05	-2.37
number of observations	427	
percent correctly predicted	73.53	
ρ^2	0.21	

Table A2 Incident Likelihood Model 2 for Crashes

Independent Variable	Estimated Coefficient	t-Statistic
constant	-0.76	-2.23
merge	0.31	1.46
visi	-0.02	-1.02
rain	1.48	3.45
number of observations	434	
percent correctly predicted	71.19	
ρ^2	0.14	